Bayesian Generation of Synthetic Streamflows

GUILLERMO J. VICENS

Resource Analysis, Inc., Cambridge, Massachusetts 02138

IGNACIO RODRÍGUEZ-ITURBE[1]

Department of Civil Engineering, Massachusetts Institute of Technology Cambridge, Massachusetts 02139

JOHN C. SCHAAKE, JR.

NOAA, Hydrologic Research Laboratory, National Weather Service Silver Spring, Maryland 20910

Generation of synthetic streamflow traces has proven to be an extremely useful technique for the evaluation of water resource planning and management alternatives. But existing models do not account for the uncertainty in the streamflow parameters. Past research efforts have focused on obtaining 'best' estimates of these parameters which are then used as the 'true' values of the process. Bayesian methods are here used to overcome this shortcoming. By incorporating the parameter uncertainties into the generation scheme, alternatives may be evaluated under both the natural and the parameter uncertainties. This is accomplished by integrating over the probability distribution of the parameters to obtain the Bayesian, predictive, or unconditional probability distribution function (pdf) of the streamflows. Use of the Bayesian pdf for synthetic generation is shown to lead, on the average, to better designs under uncertainty conditions.

Simulation models and the required synthetic streamflow generators have gained wide acceptance in water resource planning over the last decade. These models systematically investigate the effects of the variations of streamflows on the proposed development or management alternatives. *Vicens et al.* [1975] classified the uncertainties in water resources as being natural and informational, the first due to the assumed random nature of hydrologic processes and the second due to the lack of perfect information about the 'true' nature of the process.

Traditionally, research efforts have focused on each of these areas separately.. Moreover, once parameter estimation and model selection have been carried out, the remaining uncertainties about the true values of the parameters and/or the true model are ignored. Planning decisions or alternatives have been evaluated through the use of simulation models which do not account for the informational uncertainties.

The objective of this paper is to propose a Bayesian synthetic generation scheme which accounts for the parameter uncertainties due to short hydrologic record. This approach seems a more rational one in lieu of the great statistical dispersion existing in hydrologic parameters corresponding to records of common length. This paper limits itself to only one model, a first-order normal autoregressive process; other models are considered by *Vicens et al.* [1974].

Streamflow processes will be considered as random processes generating random variables distributed according to a model probability distribution function (pdf). The inherent or natural randomness of the process creates uncertainty about future observation of the process and about the consequences or benefits of any decision. In addition, the parameters of the model are unknown, a situation which further adds to the uncertainties about future observations. However, we shall assume that the model pdf is known with certainty; i.e., no model uncertainty exists.

[1] Now at Universidad Simon Bolivar, Caracas, Venezuela.

The Bayesian framework presents the possibility of including parameter uncertainties in inferences about future streamflows. First, by selecting a model we have defined a model pdf $f(y_f|\boldsymbol{\theta})$ which is the pdf of a future observation given that we know the true parameters $\boldsymbol{\theta}$. This is a conditional pdf given the parameters, but these parameters are not known. Our best information about them is described by their joint posterior pdf $f''(\boldsymbol{\theta}|I_R, \mathbf{Y})$, where I_R represents all prior regional information and $\mathbf{Y}$ represents the at-site historical record [see *Vicens et al.*, 1975]. To obtain the unconditional, marginal, predictive, or as we shall call it, the Bayesian pdf of a future streamflow, an integration over the product of the conditional pdf of y_f given $\boldsymbol{\theta}$ and the pdf of $\boldsymbol{\theta}$ is required, i.e.,

$$\tilde{f}(y_f \mid I_R, \mathbf{Y}) = \int_{\boldsymbol{\theta}} f(y_f \mid \boldsymbol{\theta}) \cdot f''(\boldsymbol{\theta} \mid I_R, \mathbf{Y}) \, d\boldsymbol{\theta} \qquad (1)$$

The resulting pdf includes the natural or inherent randomness of the streamflow process and the uncertainty about the parameters. It can be viewed as a weighted average of the model pdf with the posterior pdf of the parameters as the weights.

The Bayesian pdf can then be used for inferences about future streamflows; thus the probability of a flood over a certain flow q can be obtained by

$$P[y_f \geq q] = \int_q^{\infty} \tilde{f}(y_f \mid I_R, \mathbf{Y}) \, dy_f \qquad (2)$$

Similarly, the Bayesian moments of a future observation are given by

$$E[y_f] = \int_{y_f} y_f \cdot \tilde{f}(y_f \mid I_R, \mathbf{Y}) \, dy_f \qquad (3)$$

$$V[y_f] = \int_{y_f} (y_f - E[y_f])^2 \cdot \tilde{f}(y_f \mid I_R, \mathbf{Y}) \, dy_f \qquad (4)$$

OPTIMAL DESIGNS

Water resource planning involves the formulation and evaluation of alternative designs. Presently, deterministic optimization models are used to screen the alternatives and select a small number to be further analyzed [see *Jacoby and Loucks*, 1972]. Simulation models are then used to evaluate this smaller set of alternatives, specifically their response to the natural uncertainties of the streamflows without explicit consideration of parameter uncertainties. Both uncertainties can be brought into play by using the Bayesian pdf of future streamflows in the generation of synthetic records for the simulation run.

To evaluate any design, a value or utility function which reflects the decision maker's preferences for the outcomes must be specified. This function specifies a value for every combination of design $\mathbf{D}$ and random outcome (streamflow) y_f. A net benefit function in monetary units will be assumed to be adequate in this work. Using such a function $u(\mathbf{D}, y_f)$, the expected net benefits of any design are

$$E[u(\mathbf{D})] = \int_{y_f} u(\mathbf{D}, y_f) \cdot \tilde{f}(y_f \mid I_R, \mathbf{Y})\, dy_f \qquad (5)$$

The optimal design $\mathbf{D}^*$ is that which has the highest expected net benefits, i.e.,

$$\mathbf{D}^* \leftarrow \max_{\text{all } \mathbf{D}} E[u(\mathbf{D})] \qquad (6)$$

A design was defined as a vector quantity, since for most water resource problems, several variables need to be specified for a project. In fact, decisions are generally so complex and the relation between $\mathbf{D}$ and y_f and the net benefits is so indirect that the integral of (5) cannot be carried out analytically. This problem forces water resource planners to use simulation models for the evaluation of alternatives.

But at present the evaluation of designs is not carried out through simulations as in (5). To explain this point, that integral may be divided into two parts:

$$E[u(\mathbf{D}) \mid \boldsymbol{\theta}] = \int_{y_f} u(\mathbf{D}, y_f) \cdot f(y_f \mid \boldsymbol{\theta})\, dy_f \qquad (7)$$

and

$$E[u(\mathbf{D})] = \int_{\boldsymbol{\theta}} E[u(\mathbf{D}) \mid \boldsymbol{\theta}] \cdot f''(\boldsymbol{\theta} \mid I_R, \mathbf{Y})\, d\boldsymbol{\theta} \qquad (8)$$

Equation (7) computes the expected net benefits of a design $\mathbf{D}$ given the parameter set $\boldsymbol{\theta}$. Equation (8) then weights the expected net benefits given $\boldsymbol{\theta}$ by the joint posterior probability that $\boldsymbol{\theta}$ are the true values of the parameters. The results are the expected net benefits of $\mathbf{D}$ evaluated under both natural and parameter uncertainties. In other words, use of the Bayesian pdf for evaluation of designs is equivalent to a two-step approach which first assesses the consequences of the natural uncertainties and then the effect of the parameter uncertainty. For complex decisions, these integrals may be approximated by simulation models.

Present day simulation models ignore the parameter uncertainty. In effect, only the first integral of the two-step approach (equation (7)) is carried out. This approach is therefore not as complete as the Bayesian methodology which carries out a more complete analysis of the uncertainties involved in the design problems. The two-step approach is practically inferior to using the Bayesian pdf because only one integration needs to be approximated through simulation (equation (5)), since the Bayesian pdf (equation (1)) can generally be obtained through direct integration.

As an example of the Bayesian approach to synthetic streamflow generation we will analyze the first-order normal autoregressive process.

FIRST-ORDER NORMAL AUTOREGRESSIVE PROCESS

A time series of annual streamflows $\mathbf{Y}^t = [y_1, y_2, \cdots, y_n]$ is assumed to have been generated by a first-order autoregressive process:

$$y_t = \beta_1 + \beta_2 y_{t-1} + \epsilon_t \qquad t = 1, 2, \cdots, n \qquad (9)$$

where ϵ_t are independent and identically distributed normal random disturbances, with mean 0 and variance σ_ϵ^2. There are three unknown parameters in this model: β_1, β_2, and σ_ϵ^2. Our objective is to derive the Bayesian or predictive pdf for this model. In addition, regional and historical data will be combined to make optimum use of all available information. The more traditional version of this model often seen in the hydrologic literature is

$$(y_t - \mu) = \rho(y_{t-1} - \mu) + \sigma(1 - \rho^2)^{1/2} v_t \qquad (10)$$

where v_t is a normal disturbance term with mean 0 and variance 1. The parameters μ, σ^2, and ρ are the mean, variance, and serial correlation coefficient of the annual streamflow series. The model of (10) reduces to (9) if

$$\beta_1 = \mu(1 - \rho) \qquad (11)$$

$$\beta_2 = \rho \qquad (12)$$

$$\sigma_\epsilon = \sigma(1 - \rho^2)^{1/2} \qquad (13)$$

Equations (11)–(13) show the relation between the parameters of both models when they are all known with certainty. When parameter uncertainty exists, care must be taken in transferring information from one set of parameters to the other. In a Bayesian context these parameters are assumed to be random variables and have pdf's which contain the information available. Derived distribution theory, or at least first-order analysis [*Cornell*, 1972], must be used in transferring information from one parameter set to the other.

One difference between the traditional model and the approach described in this paper is that the process will not necessarily be limited to stationary processes (i.e., $|\beta_2| \leq 1$). The parameter β_2 will be considered a random variable with limits $-\infty < \beta_2 < \infty$; nevertheless the application of the model to observed hydrologic time series generally reduces the high probability region of β_2 to between -1 and $+1$.

Likelihood of the sample. The distribution of the disturbance term ϵ_t as assumed in the previous section is a normal pdf given by

$$f_N(\epsilon_t \mid 0, \sigma_\epsilon^2) = (2\pi)^{-1/2} \sigma_\epsilon^{-1} \cdot \exp\left[-\frac{1}{2\sigma_\epsilon^2}(\epsilon_t - 0)^2\right] \qquad (14)$$

The model pdf of y_t is then

$$f(y_t | \beta_1, \beta_2, \sigma_\epsilon^2, y_{t-1}) = f_N(y_t | \beta_1 + \beta_2 y_{t-1}, \sigma_\epsilon^2) \qquad (15)$$

i.e., when β_1, β_2, σ_ϵ^2 and y_{t-1} are known, the pdf of y_t is normal with mean and variance:

$$E[y_t \mid \beta_1, \beta_2, \sigma_\epsilon, y_{t-1}] = \beta_1 + \beta_2 y_{t-1} \qquad (16)$$

$$V[y_t \mid \beta_1, \beta_2, \sigma_\epsilon, y_{t-1}] = \sigma_\epsilon^2 \qquad (17)$$

The likelihood function for the sample $\mathbf{Y}^t = [y_1, y_2, \cdots, y_n]$ is the product of the pdf's for the individual y_t and is given by

$$L(\beta, \sigma_\epsilon \mid \mathbf{Y}, y_0) \propto \frac{1}{\sigma_\epsilon^n} \exp\left[-\frac{1}{2\sigma_\epsilon^2} \cdot \left(\sum_{t=1}^{n} (y_t - \beta_1 - \beta_2 y_{t-1})\right)^2\right] \quad (18)$$

where an initial observation y_0 is assumed to be known. This expression can be simplified if the sample statistics are defined as

$$\mathsf{V}^{-1} = \begin{bmatrix} n & \sum_{t=1}^{n} y_{t-1} \\ \sum_{t=1}^{n} y_{t-1} & \sum_{t=1}^{n} y_{t-1}^2 \end{bmatrix} \quad (19)$$

$$\mathbf{b} = \mathsf{V} \cdot \begin{bmatrix} \sum_{t=1}^{n} y_t \\ \sum_{t=1}^{n} y_{t-1} y_t \end{bmatrix} \quad (20)$$

$$\nu = n - 2 \quad (21)$$

$$s^2 = \frac{1}{\nu}\left[\sum_{t=1}^{n} (y_t - b_1 - b_2 y_{t-1})^2\right] \quad (22)$$

Then the likelihood function may be written as

$$L(\boldsymbol{\beta}, \sigma_\epsilon \mid \mathbf{Y}, y_0) \propto \frac{1}{\sigma_\epsilon^n} \cdot \exp\left\{\frac{1}{2\sigma_\epsilon^2}[\nu s^2 + (\boldsymbol{\beta} - \mathbf{b})^t \mathsf{V}^{-1}(\boldsymbol{\beta} - \mathbf{b})]\right\} \quad (23)$$

which has a kernel similar to the product of a bivariate normal and an inverted gamma 2 pdf [*Vicens et al.*, 1974].

The historical record for the Blackwater River at Webster, New Hampshire (U.S. Geological Survey station 1-870), will be used as an example. This record was broken into three samples (n = 10, 20, 42), and the sufficient statistics computed. The values shown in the first paragraph of the appendix will be used later in this paper.

Prior pdf for $\boldsymbol{\beta}$ *and* σ_ϵ. The prior information about $\boldsymbol{\beta}$ and σ_ϵ is included through their joint prior pdf. The functional form of the joint prior pdf is selected from the natural conjugate family of the likelihood function. In this case it is the product of a bivariate normal for $\boldsymbol{\beta}$ given σ_ϵ, and an inverted gamma 2 for σ_ϵ, i.e.,

$$f'(\boldsymbol{\beta}, \sigma_\epsilon \mid I_R) = f_N'^{(2)}(\boldsymbol{\beta} \mid \mathbf{b}', \sigma_\epsilon^2 \mathsf{V}') \cdot f_{IG2}'(\sigma_\epsilon \mid s', \nu') \quad (24)$$

where the prior parameters are $\mathbf{b}'$, V', s', and ν'.

The marginal pdf for $\boldsymbol{\beta}$ is obtained by integrating over σ_ϵ:

$$f(\boldsymbol{\beta} \mid I_R) = \int_0^\infty f'(\boldsymbol{\beta}, \sigma_\epsilon \mid I_R)\, d\sigma_\epsilon = f_S^{(2)}(\boldsymbol{\beta} \mid \mathbf{b}', s'^2\mathsf{V}', \nu') \quad (25)$$

which is a bivariate Student t with parameters $\mathbf{b}'$, $s'^2\mathsf{V}'$, and ν'. Further, the marginal pdf's for β_1 and β_2 are each of the univariate Student t form with moments

$$E[\beta_1 \mid I_R] = b_1' \qquad V[\beta_1 \mid I_R] = \frac{\nu'}{\nu' - 2} \cdot s'^2 \cdot v_{11}' \qquad \nu' > 2 \quad (26)$$

$$E[\beta_2 \mid I_R] = b_2' \qquad V[\beta_2 \mid I_R] = \frac{\nu'}{\nu' - 2} s'^2 \cdot v_{22}' \qquad \nu' > 2 \quad (27)$$

$$\mathrm{Cov}\,[\beta_1, \beta_2 \mid I_R] = \frac{\nu'}{\nu' - 2} s'^2 v_{12}' \qquad \nu' > 2 \quad (28)$$

where

$$\mathbf{b}' = \begin{bmatrix} b_1' \\ b_2' \end{bmatrix} \qquad \mathsf{V}' = \begin{bmatrix} v_{11}' & v_{12}' \\ v_{21}' & v_{22}' \end{bmatrix} \quad (29)$$

In a similar fashion, the marginal pdf of σ_ϵ may be obtained by

$$f'(\sigma_\epsilon \mid I_R) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f'(\boldsymbol{\beta}, \sigma_\epsilon \mid I_R)\, d\boldsymbol{\beta} = f_{IG2}'(\sigma_\epsilon \mid s', \nu') \quad (30)$$

i.e., an inverted gamma 2 with parameters s' and ν', with moments

$$E[\sigma_\epsilon \mid I_R] = \left(\frac{\nu'}{2}\right)^{1/2} s' \frac{\Gamma[(\nu' - 1)/2]}{\Gamma[\nu' - 2]} \qquad \nu' > 1 \quad (31)$$

$$V[\sigma_\epsilon \mid I_R] = \frac{\nu' s'^2}{\nu' - 2} - [E[\sigma \mid I_R]]^2 \qquad \nu' > 2 \quad (32)$$

By derived distribution theory we may obtain the marginal prior pdf of σ_ϵ^2 which is inverted gamma 1 with parameters $\frac{1}{2}\nu'$ and $\frac{1}{2}\nu' s'^2$ and moments

$$E[\sigma_\epsilon^2 \mid I_R] = \frac{\nu'}{\nu' - 2} s'^2 \qquad \nu' > 2 \quad (33)$$

$$V[\sigma_\epsilon^2 \mid I_R] = \frac{2(\nu' s'^2)^2}{(\nu' - 2)^2(\nu' - 4)} \qquad \nu' > 4 \quad (34)$$

To define the above prior pdf's uniquely, only the parameters $\mathbf{b}'$, V', s', and ν' need to be evaluated. One technique for defining these prior parameters is to specify values for the moments of the marginal pdf's and solve equations (26), (27), (28), (33), and (34) for the prior parameters. Generally, more information is available about μ, σ^2, and ρ than about β_1, β_2, and σ_ϵ^2. If this is the case, then the following relations may be used to obtain the moments of β_1, β_2, and σ_ϵ from the moments of μ, σ, and ρ:

$$E[\beta_1 \mid I_R] \simeq E[\mu \mid I_R] \cdot [1 - E[\rho \mid I_R]] \quad (35)$$

$$V[\beta_1 \mid I_R] \simeq [1 - E[\rho \mid I_R]]^2 \cdot V[\mu \mid I_R] + E^2[\mu \mid I_R] \cdot V[\rho \mid I_R] \quad (36)$$

$$E[\beta_2 \mid I_R] = E[\rho \mid I_R] \quad (37)$$

$$V[\beta_2 \mid I_R] = V[\rho \mid I_R] \quad (38)$$

$$\mathrm{Cov}\,[\beta_1, \beta_2 \mid I_R] = -\tfrac{1}{2}E[\mu \mid I_R] \cdot V[\rho \mid I_R] \quad (39)$$

$$E[\sigma_\epsilon^2 \mid I_R] \simeq E[\sigma^2 \mid I_R] \cdot [1 - E^2[\rho \mid I_R] - V[\rho \mid I_R]] \quad (40)$$

$$V[\sigma_\epsilon^2 \mid I_R] \simeq [1 - E^2[\rho \mid I_R]]^2 \cdot V[\sigma^2 \mid I_R] + 4 \cdot [E[\sigma^2 \mid I_R] \cdot E[\rho \mid I_R]]^2 \cdot V[\rho \mid I_R] \quad (41)$$

These relations were derived by using second-order approximations for the expected value of a function of a random variable and first-order approximations for the variance [see *Benjamin and Cornell*, 1970, p. 184].

As an example of how regional information can be used to arrive at prior pdf's, the Blackwater River near Webster, New Hampshire, was studied. Regression models described by *Vicens et al.* [1974] were used to predict the mean and variance of the annual flows in this river from physiographic information about its catchment. These regression relations were

TABLE 1. Prior to Posterior Analysis, Blackwater River

Information	Moments					
	$E[\beta_1]$, ft³/s	$V[\beta_1]$	$E[\beta_2]$	$V[\beta_2]$	$E[\sigma_\epsilon^2]$, ft⁶/s²	$V[\sigma_\epsilon^2]$
Prior only						
Regression models	176.5	1260.0	0.220	0.018	3046.	8.06×10^5
Sample only at						
$n = 10$	166.4	8088.	0.235	0.172	3308.	54.72×10^5
$n = 20$	137.5	2560.	0.351	0.055	2051.	6.01×10^5
$n = 42$	157.0	1157.	0.243	0.026	3112.	5.38×10^5
Posterior						
Prior and sample at						
$n = 10$	172.1	712.	0.214	0.018	2745.	4.57×10^5
$n = 20$	164.4	544.	0.230	0.010	2442.	2.77×10^5
$n = 42$	163.0	444.	0.218	0.009	2910.	2.61×10^5

The prior information is taken from the regression models in paragraph 2 of the appendix.

derived from an analysis of 106 other New England rivers. The predictions, $E[\mu]$ and $E[\sigma^2]$, and their errors, $V[\mu]$ and $V[\sigma^2]$, are presented in the second paragraph of the appendix. In addition, the average of the serial correlation coefficient for the 106 basins and the variance of these coefficients were also used, $E[\rho]$ and $V[\rho]$. Use of these values in equations (35)–(41) yielded the moments of β_1, β_2, and σ_ϵ^2. Finally, the moments of these variables were used in equations (26)–(34) to solve for the prior parameters $\mathbf{b}'$, $\mathbf{V}'$, s', and ν'. From paragraph 2 of the appendix it can be shown that the standard errors of estimate for β_1, β_2, and σ_ϵ^2 are 10, 28, and 62%, respectively.

A second prior pdf was derived from subjective assessments. A simple version of the Thomas model described by *Fiering* [1967, p. 69] was used to arrive at the predictions for μ, σ^2, and ρ. This model, described by *Vicens et al.* [1974, 1975], uses information about the precipitation process and the hydrologist's judgments about the percentage losses to estimate the moments of the streamflow parameters. The results of this analysis are shown in the third paragraph of the appendix.

Prior to posterior analysis. Prior and sample information are combined to obtain the posterior pdf of the parameters:

$$f''(\boldsymbol{\beta}, \sigma_\epsilon | I_R, \mathbf{Y}, y_0) \propto f'(\boldsymbol{\beta}, \sigma_\epsilon | I_R) \cdot L(\boldsymbol{\beta}, \sigma_\epsilon | \mathbf{Y}, y_0) \quad (42)$$

Since the prior pdf was selected from the natural conjugate family of the likelihood function, the posterior pdf is of the same form, the product of a bivariate normal and an inverted gamma 2:

$$f''(\beta, \sigma_\epsilon | I_R, \mathbf{Y}, y_0) = f_{NIG2}''(\boldsymbol{\beta}, \sigma_\epsilon | \mathbf{b}'', \mathbf{V}'', s''^2, \nu'') \quad (43)$$

with posterior parameters:

$$\mathbf{V}''^{-1} = \mathbf{V}'^{-1} + \mathbf{V}^{-1} \quad (44)$$

$$\mathbf{b}'' = \mathbf{V}'' [\mathbf{V}'^{-1} \cdot \mathbf{b}' + \mathbf{V}^{-1} \cdot \mathbf{b}] \quad (45)$$

$$\nu'' = \nu' + \nu + 2 \qquad \nu' > 0 \qquad \nu > 0 \quad (46)$$

$$s''^2 = \frac{1}{\nu''} [\nu' s'^2 + \mathbf{b}'^t \mathbf{V}'^{-1} \mathbf{b}' + \nu s^2 + \mathbf{b}^t \mathbf{V}^{-1} \mathbf{b} - \mathbf{b}''^t \mathbf{V}''^{-1} \mathbf{b}''] \quad (47)$$

The marginal posterior pdf's are of the same form as the marginal priors, with the posterior parameters replacing the prior parameters.

If no sample exists, the posterior parameters are identical to the prior parameters. If no prior information is available, the posterior parameters are identical to those of the sample set.

The prior information derived in the previous section was combined with the historical record for the Blackwater River at Webster, New Hampshire. These three series (paragraph 1 in the appendix) were combined with the two prior pdf's from the regression models and from the subjective model. Tables 1 and 2 show the marginal moments of β_1, β_2, and σ_ϵ^2 for prior, sample, and posterior information.

TABLE 2. Prior to Posterior Analysis, Blackwater River

Information	Moments					
	$E[\beta_1]$, ft³/s²	$V[\beta_1]$	$E[\beta_2]$	$V[\beta_2]$	$E[\sigma_\epsilon^2]$, ft⁶/s²	$V[\sigma_\epsilon^2]$
Prior only						
Subjective models	199.2	2385.	0.220	0.018	2498.	62.39×10^5
Sample only at						
$n = 10$	166.4	8088.	0.235	0.172	3308.	54.72×10^5
$n = 20$	137.5	2560.	0.351	0.055	2051.	6.01×10^5
$n = 42$	157.0	1157.	0.243	0.026	3112.	5.38×10^5
Posterior						
Prior and sample at						
$n = 10$	174.5	771.	0.208	0.014	2184.	7.95×10^5
$n = 20$	164.1	536.	0.233	0.010	1853.	3.12×10^5
$n = 42$	162.5	511.	0.220	0.011	2817.	3.60×10^5

The prior information is taken from the subjective assessment in paragraph 3 of the appendix.

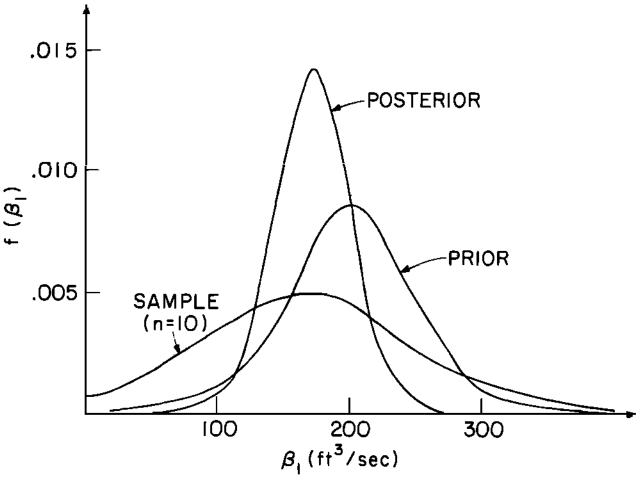

Fig. 1. Marginal pdf of β_1 (sample: n = 10).

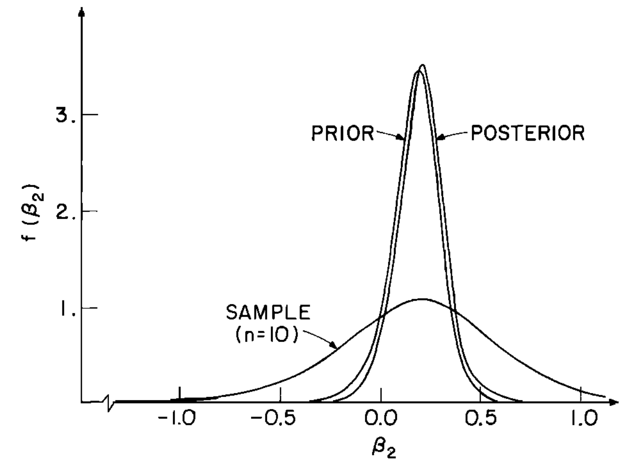

Fig. 3. Marginal pdf of β_2 (sample: n = 10).

These results show that the posterior information is, as expected, a combination of the prior and sample information. The longer the sample record, the more its influence is reflected in the posterior results. In addition, the variance of any of the three unknown parameters is always lower when the prior and sample are combined (into the posterior) than when they are used separately.

These observations are demonstrated by Figures 1–6, which show the marginal prior, sample, and posterior pdf's for two sample lengths. The prior used in all the figures was the subjective prior (paragraph 3 in the appendix). These figures show that the posterior is a combination of the prior and sample information. In addition, parameter uncertainty is reduced by 'pooling' the two sources of information.

It is interesting to point out in Figure 3 that such a small sample (n = 10) is not enough to strictly limit β_2 to between -1 and $+1$. It is possible, with a small probability, that $|\beta_2| > 1$. The prior pdf has more information and has limited β_2 to $-0.4 \leq \beta_2 \leq 0.6$. As the sample grows, in Figure 4, the sample and posterior marginal pdf's of β_2 are concentrated in the region $-0.1 \leq \beta_2 \leq 0.5$.

Bayesian distribution. More important than parameter estimates is the Bayesian or predictive distribution of the process. Whether used for inferences or decision making this pdf contains the natural uncertainty from the model distribution of the streamflows and the uncertainty about the parameters due to a lack of sufficient information.

Previously, the model pdf of a future observation was assumed to be normal. The Bayesian pdf integrates the model pdf over the joint of the unknown parameters:

$$\tilde{f}(y_f \mid I_R, \mathbf{Y}, y_0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y_t \mid \boldsymbol{\beta}, \sigma_\epsilon) \cdot f''(\boldsymbol{\beta}, \sigma_\epsilon \mid I_R, \mathbf{Y}, y_0)\, d\boldsymbol{\beta}\, d\sigma_\epsilon = \tilde{f}_s(y_f \mid m, \psi, \nu) \tag{48}$$

a Student t pdf with parameters

$$m = \mathbf{Z}^t \cdot \mathbf{b}'' \tag{49}$$

$$\psi = s''^2[1 + \mathbf{Z}^t \cdot \mathbf{V}'' \cdot \mathbf{Z}] \tag{50}$$

and

$$\xi = \nu'' \tag{51}$$

where

$$\mathbf{Z} = \begin{bmatrix} 1 \\ y_n \end{bmatrix} \tag{52}$$

and y_n is the last observation. This pdf has moments

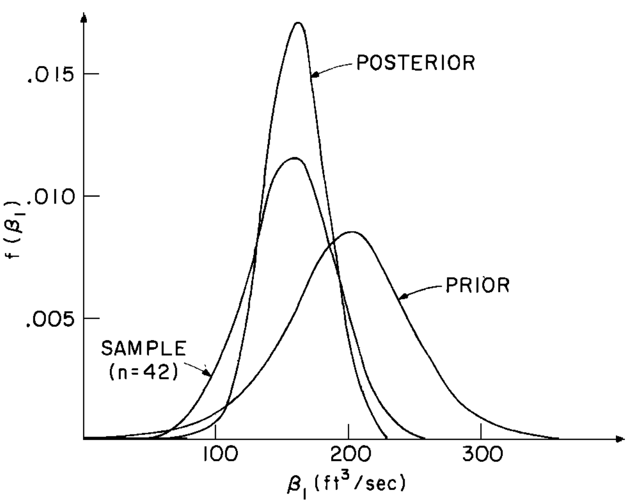

Fig. 2. Marginal pdf of β_1 (sample: n = 42).

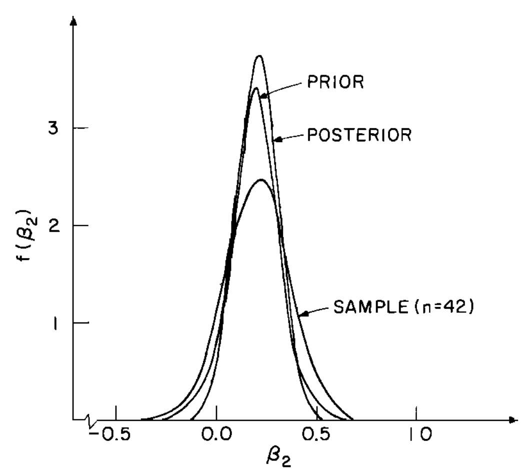

Fig. 4. Marginal pdf of β_2 (sample: n = 42).

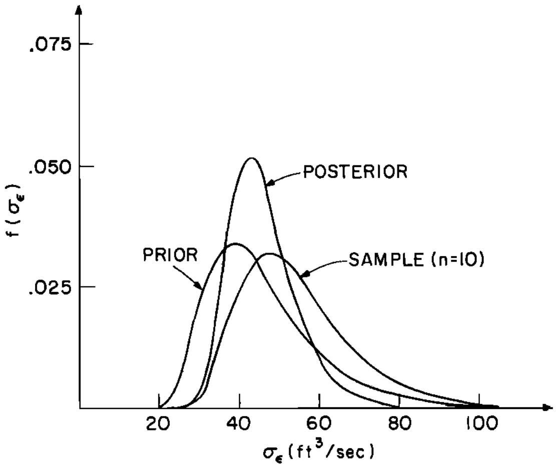

Fig. 5. Marginal pdf of σ_ϵ (sample: n = 10).

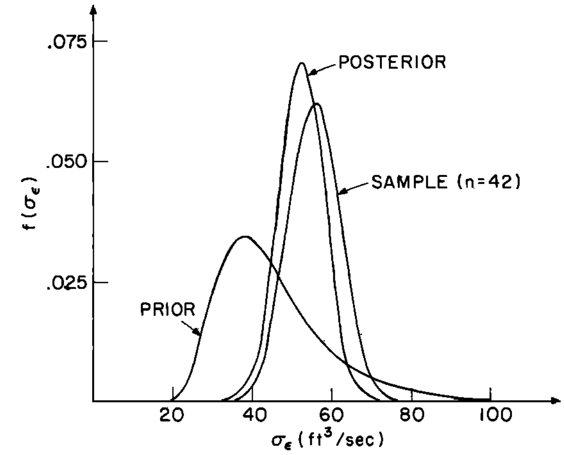

Fig. 6. Marginal pdf of σ_ϵ (sample: n = 42).

$$E[y_f|I_R, \mathbf{Y}] = m = b_1'' + b_2''y_n \tag{53}$$

$$V[y_f \mid I_R, \mathbf{Y}] = \frac{\xi\psi}{\xi - 2} = \frac{\nu''}{\nu'' - 2}\cdot s''^2[1 + \mathbf{Z}^t\mathbf{V}''\mathbf{Z}] \tag{54}$$

Use of the Bayesian pdf for the generation of synthetic streamflow will produce traces which may differ in their statistics from those of the historical record. This is a consequence of (1) using more than one source of information for data about the unknown parameters and (2) explicitly including the parameter uncertainty in the distribution of the streamflows by integrating the product of the model pdf (natural uncertainty) and the joint pdf of the parameters. In general, the most significant difference between the synthetic traces and the historical record will be a larger streamflow variance due to the parameter uncertainty. In addition, there may be differences in the mean or serial correlation if an informative prior pdf is used.

In generating synthetic traces for simulation purposes, the first streamflow of every trace should be generated from equation (48). The second sample should be generated from a similar distribution except that y_f, the first generated sample, replaces y_n in equation (52). All of the future flows should be generated in a similar manner, by substitution of the last sample generated into equation (32). Therefore the parameter uncertainty is accounted for in the entire trace.

The Bayesian pdf was derived for all combinations of the three samples and the two prior pdf's. These results are shown in Tables 3 and 4. The posterior expected values of y_f are not necessarily weighted averages of the prior and the sample means. The future streamflow is dependent on the last observation as shown by equation (48). This is why the Bayesian pdf is also called a predictive pdf. The moments in Tables 3 and 4 only apply to the next observation.

The posterior variance of y_f is generally lower than the prior or sample variances. This is a direct result of the reduction of the parameter uncertainty. Variations occur, since $V[y_f]$ is also a function of the last observation.

BAYESIAN APPROACH FOR WATER RESOURCE DECISION MAKING

Decision problem: reservoir design. The decision problem will be the following: select a reservoir capacity and target release which will maximize the expected discounted net benefits of the project over the next N years. The design purpose is to meet an annual water supply demand. The reservoir costs will in this example be given by the cost function:

$$C(S) = C_1 + C_2S \tag{55}$$

where S is the capacity and C_1 and C_2 are coefficients, whereas the annual net benefits of a release R, when the target release was T, are

$$\begin{aligned} B(R, T) &= B_1 + \lambda T - \delta(T - R) \qquad T \geq R \\ B(R, T) &= B_1 + \lambda T + \gamma(R - T) \qquad R \geq T \end{aligned} \tag{56}$$

when T and R are in cubic feet. This function is shown by Figure 7. This benefit function contains an opportunity loss for not obtaining the planned for target level or for underestimating and not planning for a higher level. These short-run losses will penalize designs which are too conservative or too optimistic about the achievable yield from the stream.

The discounted net benefits for any design $\mathbf{D} = [T, S]$ are

$$NB(\mathbf{D}) = \sum_{t=0}^{N} [B(R, T)/(1 + r)^t] - C_1 - C_2S \tag{57}$$

where r is the discount rate, and it is assumed that the reservoir is constructed entirely in the first year. If any maintenance costs need to be accounted for, these will be included as a negative benefit in equation (56) through B_1.

A very simple optimizing criterion has been selected to simplify the design selection procedure described in the next section. Other more complex criteria may be proposed for

TABLE 3. Moments of the Bayesian pdf, Blackwater River

Information	$E[y_f]$, ft^3/s	$V[y_f]$, ft^6/s^2
Prior only		
Regression models	226.2	4455.
Sample only at		
n = 10	226.4	3953.
n = 20	212.3	2154.
n = 42	222.8	3295.
Posterior		
Prior and sample at		
n = 10	226.8	3017.
n = 20	213.5	2551.
n = 42	222.0	3023.

The prior information is taken from the regression models in paragraph 2 of the appendix.

TABLE 4. Moments of Bayesian pdf, Blackwater River

Information	$E[y_t]$, ft³/s	$V[y_t]$, ft⁶/s²
Prior only		
Subjective models	255.4	5460.
Sample only at		
$n = 10$	226.4	3953.
$n = 20$	212.3	2154.
$n = 42$	222.8	3295.
Posterior		
Prior and sample at		
$n = 10$	227.7	2418.
$n = 20$	213.7	1941.
$n = 40$	222.2	2931.

The prior information is taken from the subjective models in paragraph 3 of the appendix.

realistic cases, but this simple one is adequate for the example.

Since the future releases from any design are random variables, the net benefits of that design are also a random variable. To assess the expected net benefits, a simulation model is used.

Design procedure. To select an optimal design, a simulation search procedure will be used. This approach would be costly for complex problems of many design variables, but for this problem of only two variables, capacity and target, it is a useful and simple procedure. More important, a simulation procedure analyzes each alternative design under the uncertainty conditions arising from the random inputs, namely, streamflows. The synthetic generation of streamflows will be discussed in the next section.

To obtain an optimal design, the method of steepest ascent will be applied [*Maass et al.,* 1962, p. 399]. This iterative technique searches for the maximum of the net benefit 'surface' by moving in the direction of steepest slope of the surface from the point that it was previously at. For two design variables the procedure is the following.

1. Select a starting point, $\mathbf{D}_0 = [C_0, T_0]$.

2. Evaluate the expected net benefits of this decision through the simulation.

3. Evaluate two other alternatives,

$$\mathbf{D}' = [C_0 + \Delta C, T_0] \tag{58}$$

and

$$\mathbf{D}'' = [C_0, T_0 + \Delta T] \tag{59}$$

where ΔC and ΔT are increments in capacity and target, respectively.

4. Estimate the partial derivative of the net benefit surface through

$$\frac{\partial NB}{\partial C} = \frac{B(\mathbf{D}') - B(\mathbf{D}_0)}{\Delta C} \tag{60}$$

$$\frac{\partial NB}{\partial T} = \frac{B(\mathbf{D}'') - B(\mathbf{D}_0)}{\Delta T} \tag{61}$$

5. Move to a new point $\mathbf{D}_1 = [C_1, T_1]$, where

$$C_1 = C_0 + K\frac{\partial NB}{\partial C} \tag{62}$$

$$T_1 = T_0 + K\frac{\partial NB}{\partial T} \tag{63}$$

and

$$K = d\left[\left(\frac{\partial NB}{\partial C}\right)^2 + \left(\frac{\partial NB}{\partial T}\right)^2\right]^{-1} \tag{64}$$

The parameter d is the maximum 'distance' to be allowed in each move.

6. Evaluate the new alternative $\mathbf{D}_1$. If the expected net benefits are lower than for $\mathbf{D}_0$, return to step 5 and use a new value of d equal to one-half the previous value. If the expected net benefits are higher than those for $\mathbf{D}_0$, but only by less than ω, stop and accept D_1 as the optimal design. Otherwise return to step 3 and repeat the procedure to obtain a new design $\mathbf{D}_2$.

The major advantage of the steepest ascent method is that by measuring the derivative of the response surface the optimum may be approached quite rapidly. This procedure may not attain a global optimum if local optima exist. For this reason, a systematic search of the entire response surface is important at the beginning.

In step 6, d is set to one-half its previous value when the net benefits decrease, since taking too large a step has apparently 'stepped over' the maximum point on the surface. This procedure allows the user to specify a large d at the beginning of the optimization process. The process will converge to an area near the optimum quickly, then decrease its step size for a more careful search of the exact optimal solution.

The design variables T and S are measured in dimensionless form, in percent of the annual mean flow.

Synthetic streamflow generation. The philosophy and theory behind the use of synthetic streamflow traces for the analysis of hydrologic uncertainty in water resource planning was discussed extensively by *Maass et al.* [1962] and *Fiering* [1967]. The objective of this type of modeling has been the preservation of the historical parameters estimated from the existing records. The Bayesian approach described in this work attempts to include the parameter uncertainty as well as the natural uncertainty in the design process. The difference between these two approaches will result in different designs. A comparison of these designs is carried out later on. The

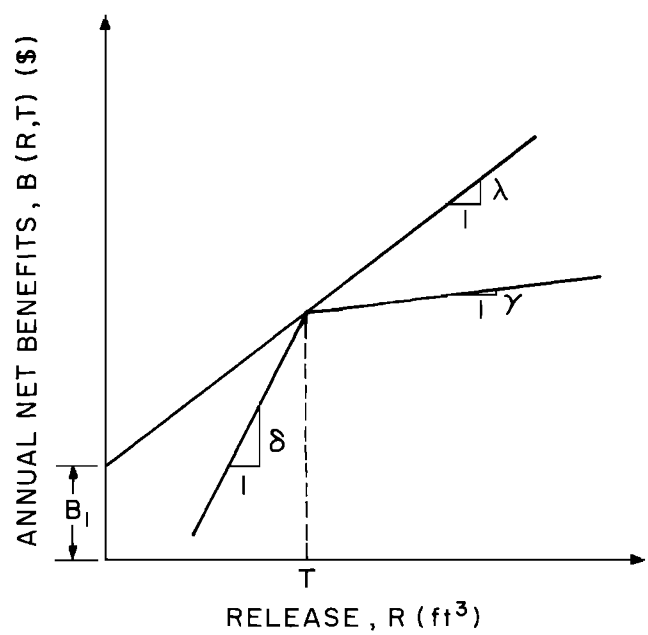

Fig. 7. Annual net benefits function.

differences between the two synthetic generating schemes are discussed first.

Model. The model used for this example is the first-order normal autoregressive process described earlier. No model uncertainty is considered; therefore it is assumed that the true model is given by

$$y_t = \beta_1 + \beta_2 y_{t-1} + \epsilon_t \quad (65)$$

where ϵ_t is a disturbance term distributed according to a normal pdf with mean 0 and variance σ_ϵ^2.

Traditional approach. As was noted before, the traditional approach to synthetic generation for this model has been to estimate the parameters of the model from the historical record and ignore the parameter uncertainty. The estimated parameters, by any of many methods which may or may not include regional data, are assumed to be the 'true' ones, and the simulation efforts are directed to the selection of an optimal design under the many possible streamflow sequences which are generated from the model.

Synthetic streamflow traces are then generated through equation (65) by using the estimates $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\sigma}_\epsilon^2$ instead of the true but unknown parameters β_1, β_2, and σ_ϵ^2. To start this procedure, the last streamflow observed in the historical record was used as a starting base.

Bayesian approach. The Bayesian approach to synthetic generation is to acknowledge the existing parameter uncertainty and explicitly include it in the analysis. This may be very important when the existing historical records are relatively short. In this very common case the uncertainty in the parameters is as important as that of the process itself. Even when relatively long historical records are available, the inherent uncertainty in a parameter such as the correlation coefficient ρ $(=\beta_2)$ is very large as shown by *Rodríguez-Iturbe* [1969]. Errors in estimating ρ of the order of 100% are common even for records of more than 50 yr.

The Bayesian pdf is therefore a powerful tool, since it includes both the natural and the parameter uncertainty. For the first-order normal autoregressive process, the Bayesian pdf of a future streamflow is

$$f(y_f|\mathbf{Y}) = \tilde{f}s(y_f|m, \psi, \xi) \quad (66)$$

a Student t pdf with parameters m, ψ, and ξ defined in equation (48). To generate synthetic traces, random numbers are generated from a Student t pdf with mean and variance as in equations (53) and (54). The streamflow generated replaces y_n in equation (53) to generate the next one.

Earlier sections of this paper have shown how several sources of information can be combined to reduce parameter uncertainties. For the purposes of comparison with a classical procedure, only the historical record information will be used here. A noninformative prior pdf will be used, as described by *Zellner* [1971, p. 187]. A so-called Jeffrey's pdf can be used to express 'ignorance':

$$f'(\boldsymbol{\beta}, \sigma_\epsilon) \propto \frac{1}{\sigma_\epsilon} \quad (67)$$

When this prior pdf is combined with the likelihood function, the posterior pdf is identical to that of equation (43) except that the posterior parameters are

$$\mathbf{V}'' = \mathbf{V} \quad (68)$$

$$\mathbf{b}'' = \mathbf{b} \quad (69)$$

$$\nu'' = \nu \quad (70)$$

$$s''^2 = s^2 \quad (71)$$

As was discussed earlier, the major differences between the traces generated by the traditional approach and the classical approach will be a lower variance in the latter, since the Bayesian approach has included the uncertainty about the parameters, which the traditional approach totally ignores.

Comparison of designs. To compare the designs obtained from using these two different synthetic generating schemes in the problem described earlier, the following experiment has been carried out.

1. Fix the true parameters equal to specific values β_1, β_2, and σ_ϵ^2.

2. Generate a ficticious historical record of length n using the true parameters.

3. Estimate the parameters in a classical framework, generate synthetic traces, and optimize the reservoir design problem, D_T^*.

4. Generate synthetic traces from the Bayesian pdf, and optimize the same design problem as before to obtain $\mathbf{D}_B^*$.

5. Evaluate designs $\mathbf{D}_T^*$ and $\mathbf{D}_B^*$ by means of simulation procedures, using synthetic sequences generated from the

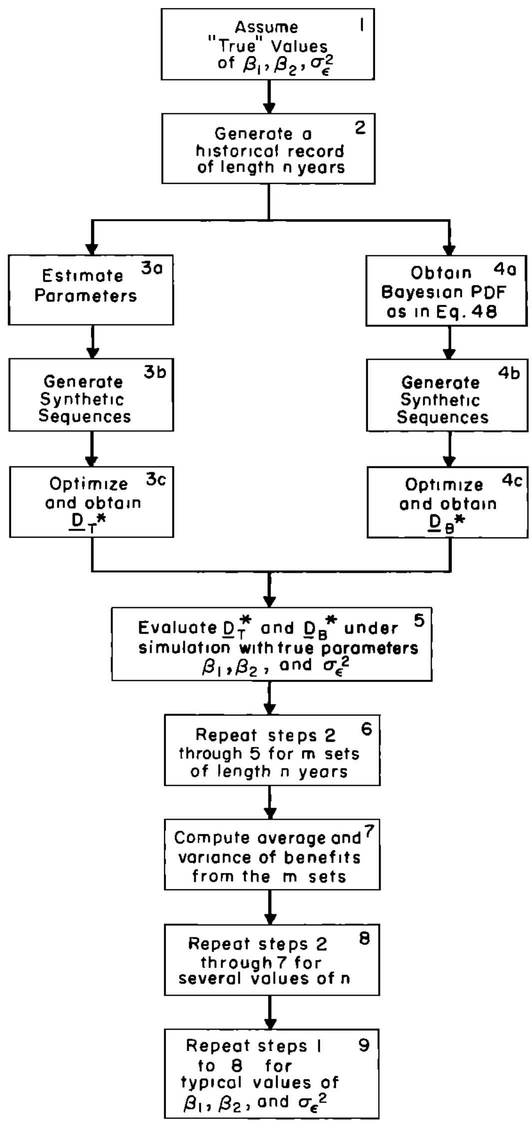

Fig. 8. Experimental procedure for comparison of designs.

true parameters, β_1, β_2, and σ_ϵ^2, to obtain $NB(\mathbf{D}_T^*|n)$ and $NB(\mathbf{D}_B^*|n)$.

6. Repeat steps 2-5 for m sets of historical records of length n.

7. Compute the average and variance of the expected net benefits for a historical record of length n, i.e., compute

$$\langle NB\rangle(\mathbf{D}_T^* \mid n) = \frac{1}{m}\sum_{i=1}^{m} NB(\mathbf{D}_{T_i}^* \mid n) \tag{72}$$

$$\hat{N}B(\mathbf{D}_T^* \mid n) = \frac{1}{m-1}\sum_{i=1}^{m} [NB(\mathbf{D}_{T_i}^* \mid n) - \langle NB\rangle(\mathbf{D}_{T_i}^* \mid n)]^2 \tag{73}$$

$$\langle NB\rangle(\mathbf{D}_B^* \mid n) = \frac{1}{m}\sum_{i=1}^{m} NB(\mathbf{D}_{B_i}^* \mid n) \tag{74}$$

$$\hat{N}B(\mathbf{D}_B^* \mid n) = \frac{1}{m-1}\sum_{i=1}^{m} [NB(\mathbf{D}_{B_i}^* \mid n) - \langle NB\rangle(\mathbf{D}_{B_i}^* \mid n)]^2 \tag{75}$$

where $\mathbf{D}_{T_i}^*$ and $\mathbf{D}_{B_i}^*$ are the optimal designs obtained in steps 3 and 4 for the ith historical records, $i = 1, 2, \cdots, m$.

8. Repeat steps 1-7 for different historical record lengths n.

9. Repeat steps 1-8 for several typical values of β_1, β_2, and σ_ϵ^2.

This entire procedure is reproduced in Figure 8.

This experiment attempts to evaluate the differences in designs and net benefits to be expected if a traditional versus a Bayesian simulation procedure is used. The average net benefits, $\langle NB\rangle(\mathbf{D}_T^*|n)$ and $\langle NB\rangle(\mathbf{D}_B^*|n)$, are estimates of the net benefits to be obtained if a traditional or Bayesian procedure was used in the future when faced with historical records of length n years. The variances, $\hat{N}B(\mathbf{D}_T^*|n)$, and $\hat{N}B(\mathbf{D}_B^*|n)$, are estimates of the variations to be expected in these future designs. A water resource planner would prefer design procedures that on the average result in higher net benefits. If in addition the procedure selected minimizes the probability of obtaining designs with very low net benefits, this procedure would be quite successful.

Results. The coefficients of the benefit and cost functions used are given below.

$$\lambda = 3.15 \times 10^4\ \$/\text{ft}^3$$
$$\gamma = 3.15 \times 10^3\ \$/\text{ft}^3$$
$$\delta = 3.15 \times 10^5\ \$/\text{ft}^3$$
$$C_1 = 0 \quad (\$)$$
$$C_2 = 200 \quad (\$/10^6\ \text{ac ft})$$

The sets of parameter values used in step 1 are in Table 5. These represent average and rare values for medium-sized streams. Parameter set 2 has a high correlation coefficient (ρ = 0.6), while set 3 has a high coefficient of variation (C_v = 0.6). Set 4 has no serial correlation at all (ρ = 0).

In step 5, 50 sets of each record length n were used. These should provide good estimates of the effect or influence of the record length n. For the synthetic sequences in steps 3*b* and 4*b* (Figure 8), 50 sets of 50 yr were generated. Again these are enough to predict the net benefits of any design accurately. The evaluation of the optimal designs $\mathbf{D}_B^*$ and $\mathbf{D}_T^*$ in step 5 was made through the use of 50 sets of 50 yr of synthetic traces generated with the true parameters.

All of the results are presented in Table 6. In addition, the average and variance of the expected net benefits are plotted in Figures 9 and 10 versus the given length of historical record for the different parameter sets. In Table 6, the third and fourth columns show the average and variance of the net benefits from traditional designs, while the seventh and eighth columns show this same value for the Bayesian design. The average optimal designs for the traditional approach are given in columns 5 and 6 ($\langle T_t^*\rangle$ and $\langle S_T^*\rangle$), and the Bayesian ones are in columns 9 and 10 ($\langle T_B^*\rangle$ and $\langle S_B^*\rangle$).

Two characteristics of the differences between the expected net benefits of the Bayesian and traditional procedures are shown. First, the average of the expected net benefits is always higher for the Bayesian approach than for the traditional approach. This difference is significant for short historical records ($n < 30$), and for high values of correlation and coefficient of variation (sets 2 and 3). Second, the variance of the expected benefits for the Bayesian approach is always much lower than it is for the traditional procedures. This indicates that over many design cases the traditional approach will design some 'wild' alternatives with very low net benefits. For example, for relatively short records ($n \sim 10$ yr) the average net benefits of the 50 Bayesian designs are 6-100% higher than the 50 traditional designs (parameter sets 4 and 3, respectively), while the variance of the Bayesian designs is 18-72% lower (parameter sets 4 and 2, respectively). Even for relatively long records (n = 50) the average net benefits of the Bayesian designs are higher and the variance is lower by more than 19%.

The difference between the net benefits of the Bayesian designs versus the traditional designs can be best explained if a specific example is considered in detail (parameter set 3 and n = 20). The average net benefit of the Bayesian designs is 31% higher, while the variance is more than 41% lower. Both procedures frequently produce designs close to the optimal one (i.e., the one which would maximize net benefits if the true parameters were known). But the traditional procedure will sometimes design reservoirs which yield very low net benefits. The left tail of a histogram of the net benefits obtained is 'larger' for the traditional approach than for the Bayesian approach. The Bayesian approach greatly reduces the probability of obtaining designs with very low net benefits.

These differences in results are due to the differences in the approach to design. The Bayesian approach acknowledges the parameter uncertainty. This leads to designs which tend to have higher storage capacities and/or lower target releases. As can be seen in Table 6, the optimal Bayesian designs $\mathbf{D}_B^*$ have lower targets and larger capacities than the traditional designs $\mathbf{D}_T^*$, especially for very short historical records ($n \leq 20$). In short, the Bayesian approach appears to be more conservative.

This does not imply that the Bayesian approach leads to better designs every time. Since both approaches rely on the historical record for their information, if the historical statistics are identical to the true parameters, the traditional approach would be 'lucky' enough to achieve the best design. But the probability of this event is very low, especially for short historical records. The more usual occurrence is for the traditional approach to either overdesign or underdesign by

TABLE 5. True Parameter Sets

Set	β_1, ft³/s	β_2	σ_3, ft³/s	μ, ft³/s	σ, ft³/s	ρ	C_v
1	1050.	0.3	429.3	1500.	450.	0.3	0.3
2	600.	0.6	360.	1500.	450.	0.6	0.3
3	1050.	0.3	858.5	1500.	900.	0.3	0.6
4	1500.	0.	450.	1500.	450.	0.	0.3

TABLE 6. Results of Traditional and Bayesian Designs

Parameter Set	Record Length, yr	$\langle NB \rangle (D_T^* \mid n)$, $\$ \times 10^6$	$\hat{N}B \cdot (D_T^* \mid n)$	$\langle D_T^* \rangle$ $\langle T_T^* \rangle$	$\langle D_T^* \rangle$ $\langle S_T^* \rangle$	$\langle NB \rangle (D_B^* \mid n)$, $\$ \times 10^6$	$\hat{N}B \cdot (D_B^* \mid n)$	$\langle D_B^* \rangle$ $\langle T_B^* \rangle$	$\langle D_B^* \rangle$ $\langle S_B^* \rangle$
1	10	373.8	42280.	0.80	0.40	407.3	19260.	0.75	0.49
	20	432.2	17270.	0.79	0.41	445.5	10410.	0.76	0.44
	30	483.3	2195.	0.76	0.40	486.2	1496.	0.74	0.42
	50	497.3	900.	0.76	0.40	499.1	723.	0.75	0.41
2	10	278.1	88630.	0.77	0.27	295.6	72160.	0.73	0.38
	20	316.7	55710.	0.75	0.25	325.6	44940.	0.73	0.35
	30	419.8	5035.	0.70	0.27	420.1	3424.	0.69	0.34
3	10	−2.4	160090.	0.61	0.55	94.6	76570.	0.54	0.61
	20	115.1	60320.	0.57	0.57	150.8	35460.	0.53	0.59
	30	203.4	6320.	0.52	0.55	212.8	3746.	0.49	0.55
4	10	443.9	26570.	0.84	0.51	474.5	7304.	0.78	0.58
	20	507.6	6789.	0.83	0.49	513.9	3339.	0.80	0.51

large values, whereas the Bayesian approach will do the same but by lower amounts.

Use of informative prior pdf's. Previous work showed that the effect of using informative prior pdf's is to reduce the parameter uncertainty, which is analogous to increasing the total equivalent sample size. On the average, i.e., over many cases, the use of an informative prior will result in designs which are 'closer' to the true optimal designs, and with less variance. This is a result of reducing the posterior variance of the parameters, which in turn leads to lower uncertainty levels in the posterior information. Again this is the case on the average, since for one particular case the prior pdf could be mistaken.

An implication of the results of the previous section is that for long historical records ($n > 40$ or 50) it would not be profitable to bother with an informative prior pdf. The marginal increase in the average net benefits of the Bayesian designs is very small.

However, for relatively short historical records ($n \sim 10$–30) which are more common, informative prior pdf's would be quite valuable. An increase in the total information from 10 to 20 'equivalent' total years could increase the expected average net benefits by roughly 50–100% in this particular example. This possibility should obviously foster research interest in this area.

Model uncertainty. An assumption was made that the true model was known. This assumption stated that the model was of the autoregressive family, and the disturbance terms were normally distributed. This assumption eliminates all model uncertainty from this problem.

The model uncertainty problem can be divided into two questions: (1) structural model uncertainty (e.g., is the process autoregressive or fractional noise?) and (2) distributional model uncertainty (e.g., are the error terms normal or log normal?). For decisions such as those presented in this paper the distributional model uncertainty is not very important. Unless

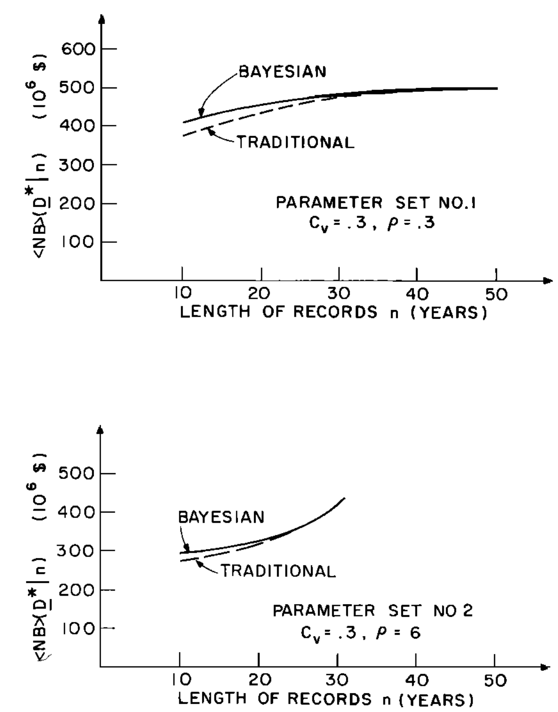

Fig. 9a. Average expected net benefits (parameter sets 1 and 2).

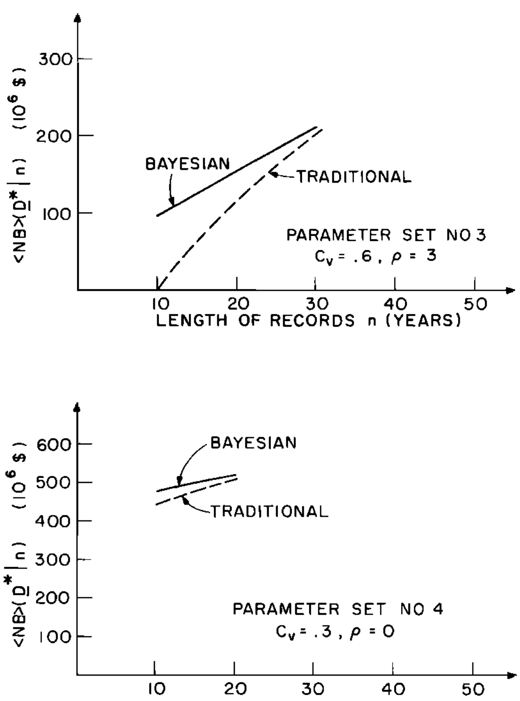

Fig. 9b. Average expected net benefits (parameter sets 3 and 4).

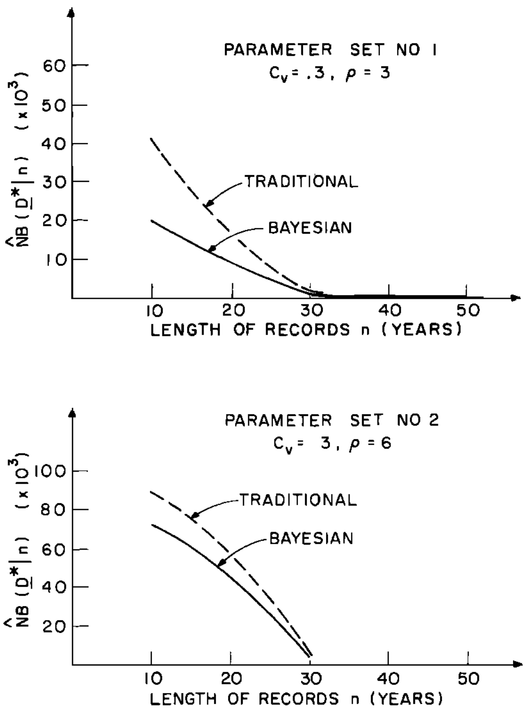

Fig. 10*a*. Variance of the expected net benefits (parameter sets 1 and 2).

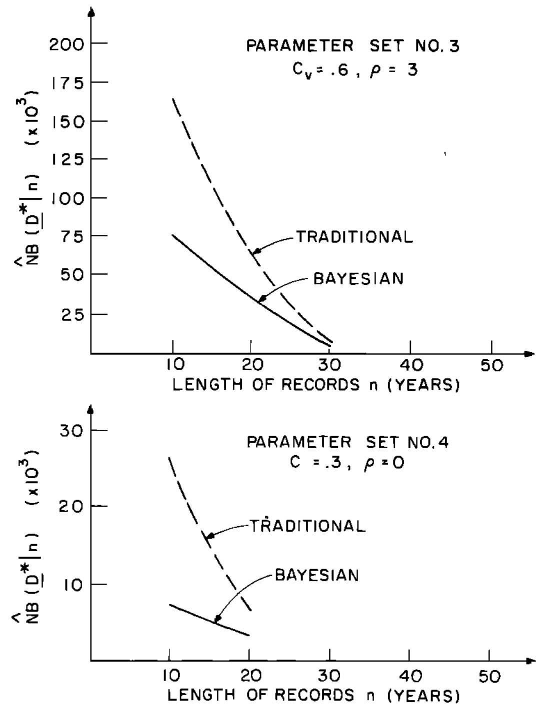

Fig. 10*b*. Variance of the expected net benefits (parameter sets 3 and 4).

extreme events are of interest, the distributional uncertainty can be ignored (see *Rodríguez-Iturbe et al.* [1971, p. 1141]). The problem of structural model uncertainty is probably more important, especially for designs that seek high levels of yield from the stream in comparison with the mean flow. This problem and solutions for it are still the subject of research. No general conclusions have been arrived at.

This section presented a comparison of Bayesian and traditional design procedures for various lengths of historical records and several sets of typical streamflow parameters. A reservoir capacity and target release decision was considered under the assumption that the true model was a first-order normal autoregressive process.

For the particular decision and economic coefficients considered, the Bayesian approach leads to 'better' designs on the average. This implies that if some agency were to select a procedure to carry out water resource designs, the use of Bayesian procedures will, on the average, lead to designs with higher net benefits. In addition, Bayesian procedures greatly reduce the probability of designs with very low net benefits.

Conclusions

A Bayesian approach to the generation of synthetic streamflow traces from a first-order normal autoregressive process was presented. This approach appears to be a more rational approach, since it explicitly accounts for the parameter uncertainties in the generation of synthetic streamflows. In addition, it allows the use of other sources of information in conjunction with the historical record to reduce the parameter uncertainties.

Appendix

1. Sample information for the Blackwater River, Webster, New Hampshire, is given below.

Sample $n = 10$ (1928–1938)

$$\mathbf{b} = \begin{bmatrix} 166.4 \\ 0.2349 \end{bmatrix} \qquad \mathbf{V}^{-1} = \begin{bmatrix} 10. & 2125. \\ 2125. & 470{,}660. \end{bmatrix}$$

$$s^2 = 2481. \qquad \nu = 8$$

Sample $n = 20$ (1928–1948)

$$\mathbf{b} = \begin{bmatrix} 137.4 \\ 0.3509 \end{bmatrix} \qquad \mathbf{V}^{-1} = \begin{bmatrix} 20. & 4235. \\ 4235. & 934{,}290. \end{bmatrix}$$

$$s^2 = 1824. \qquad \nu = 18$$

Sample $n = 42$ (1928–1970)

$$\mathbf{b} = \begin{bmatrix} 157.0 \\ 0.2431 \end{bmatrix} \qquad \mathbf{V}^{-1} = \begin{bmatrix} 42. & 8628. \\ 8628. & 1{,}893{,}870. \end{bmatrix}$$

$$s^2 = 2955. \qquad \nu = 40$$

2. Prior information from regression model and regional serial correlation coefficient for the Blackwater River, Webster, New Hampshire, are given below.

$$E[\mu | I_R] = 226.2 \text{ ft}^3/\text{s} \qquad V[\mu | I_R] = 515.$$

$$E[\sigma^2 | I_R] = 3263.5 \text{ ft}^6/\text{s}^2 \qquad V[\sigma^2 | I_R] = 822{,}050.$$

$$E[\rho | I_R] = 0.22 \qquad V[\rho | I_R] = 0.01844$$

$$E[\beta_1 | I_R] = 176.5 \text{ ft}^3/\text{s} \qquad V[\beta_1 | I_R] = 1256.9$$

$$E[\beta_2 | I_R] = 0.22 \qquad V[\beta_2 | I_R] = 0.01844$$

$$\text{Cov}\,[\beta_1, \beta_2 | I_R] = -2.09$$

$$E[\sigma_\epsilon^2 | I_R] = 3045.5 \text{ ft}^6/\text{s}^2 \qquad V[\sigma_\epsilon^2 | I_R] = 80{,}532$$

$$\mathbf{b}' = \begin{bmatrix} 176.5 \\ 0.22 \end{bmatrix} \qquad \mathbf{V}'^{-1} = \begin{bmatrix} 2.98 & 337.3 \\ 337.3 & 203{,}312. \end{bmatrix}$$

$$s' = 2819.9 \qquad \nu' = 27$$

3. Prior information from subjective assessments for the Blackwater River, Webster, New Hampshire, is given below.

$$E[\mu|I_R] = 255.4 \text{ ft}^3/\text{s} \qquad V[\mu|I_R] = 1942.6$$

$$E[\sigma^2|I_R] = 2676.5 \text{ ft}^6/\text{s}^2 \quad V[\sigma^2|I_R] = 5{,}509{,}000.$$

$$E[\rho|I_R] = 0.22 \qquad V[\rho|I_R] = 0.01844$$

$$E[\beta_1|I_R] = 199.2 \text{ ft}^3/\text{s} \qquad V[\beta_1|I_R] = 2385.0$$

$$E[\beta_2|I_R] = 0.22 \qquad V[\beta_2|I_R] = 0.01844$$

$$\text{Cov}[\beta_1, \beta_2|I_R] = -2.35$$

$$E[\sigma_\epsilon^2|I_R] = 2497.7 \text{ ft}^6/\text{s}^2 \quad V[\sigma_\epsilon^2|I_R] = 62{,}386{,}000.$$

$$\mathbf{b}' = \begin{bmatrix} 199.2 \\ 0.22 \end{bmatrix} \qquad \mathbf{V}'^{-1} = \begin{bmatrix} 1.20 & 153.0 \\ 153.0 & 154{,}993. \end{bmatrix}$$

$$s'^2 = 1665. \qquad \nu' = 6$$

NOTATION

- $\mathbf{b}'$ prior parameters.
- $\mathbf{b}$ sample parameters.
- $\mathbf{b}''$ posterior parameters.
- d design alternative.
- $\mathbf{D}$ vector of design decisions.
- $E[\ ]$ expected value operator.
- $f(\ )$ probability distribution function.
- $F(\ )$ cumulative distribution function.
- $f(y|\boldsymbol{\theta})$ conditional pdf of y given the parameter set $\boldsymbol{\theta}$.
- $\tilde{f}(y)$ Bayesian, unconditional, or predictive pdf of y.
- $f'(\boldsymbol{\theta})$ prior pdf of the parameter set $\boldsymbol{\theta}$.
- $f''(\boldsymbol{\theta})$ posterior pdf of the parameter set $\boldsymbol{\theta}$.
- I_R prior regional information.
- $k(\boldsymbol{\theta})$ kernel of the likelihood function of $\boldsymbol{\theta}$.
- $L(\boldsymbol{\theta}|\mathbf{Y})$ likelihood function of $\boldsymbol{\theta}$ given the observations $\mathbf{Y}$.
- NB net benefits.
- $\langle NB \rangle$ average of net benefits.
- $\hat{NB}$ variance of the net benefits.
- R reservoir release.
- s'^2 prior variance parameter.
- s^2 sample variance parameter.
- s''^2 posterior variance parameter.
- T reservoir target release.
- $u(\ )$ utility function.
- $U(\ )$ total utility.
- $V[\ ]$ variance operator.
- $\mathbf{V}'$ prior parameters of first-order normal autoregressive process.
- $\mathbf{V}$ sample parameters of first-order normal autoregressive process.
- $\mathbf{V}''$ posterior parameters of first-order normal autoregressive process.
- y_t observation of streamflow process.
- $\mathbf{Y}$ set of observation of y.
- y_f future streamflow.
- $\boldsymbol{\beta}$ parameters of autoregressive model.
- ϵ_t disturbance term in autoregressive models.
- $\boldsymbol{\epsilon}$ set of ϵ_t.
- $\boldsymbol{\theta}$ parameter vector.
- ν' prior degrees of freedom.
- ν sample degrees of freedom.
- ν'' posterior degrees of freedom.
- π equal to 3.1416.
- ρ serial correlation.
- σ^2 process variance.
- σ_ϵ^2 variance of disturbance term ϵ_t.
- μ process mean.

Acknowledgments. This work was performed at the Ralph M. Parsons Laboratory for Water Resources and Hydrodynamics at Massachusetts Institute of Technology as part of the first author's doctoral dissertation. The support of the Office of Water Resources Research, grant 14-31-001-9021, and the National Science Foundation, grant GK-41643X, is gratefully acknowledged. The enlightening discussions with Gordon M. Kaufman of the Alfred P. Sloan School of Management at Massachusetts Institute of Technology were very much appreciated.

REFERENCES

Benjamin, J. R., and C. A. Cornell, *Probability, Statistics, and Decision for Civil Engineers,* McGraw Hill, New York, 1970.

Cornell, C. A., First-order analysis of model and parameter uncertainties, paper presented at International Symposium on Uncertainties in Hydrology and Water Resource Systems, Univ. of Ariz., Tucson, Dec. 1972.

Emshoff, J. R., and R. L. Sisson, *Design and Use of Computer Simulation Models,* Macmillan, New York, 1970.

Fiering, M. B, *Streamflow Synthesis,* Harvard University Press, Cambridge, Mass., 1967.

Jacoby, H. D., and D. P. Loucks, Combined use of optimization and simulation models in river basin planning, *Water Resour. Res., 8*(6), 1401–1414, 1972.

Maass, A., et al., *Design of Water Resource Systems,* Harvard University Press, Cambridge, Mass., 1962.

Raiffa, H., and R. Schlaifer, *Applied Statistical Decision Theory,* MIT Press, Cambridge, Mass., 1961.

Rodríguez-Iturbe, I., Estimation of statistical parameters for annual river flows, *Water Resour. Res., 5*(6), 1418–1421, 1969.

Rodríguez-Iturbe, I., D. R. Dawdy, and L. E. Garcia, Adequacy of Markovian models with cyclic components for stochastic streamflow simulation, *Water Resour. Res., 7*(5), 1127–1143, 1971.

Vicens, G. J., I. Rodríguez-Iturbe, and J. C. Schaake, Jr., A Bayesian approach to hydrologic time series modeling, *Rep. 181,* Ralph M. Parsons Lab. for Water Resour. and Hydrodyn., Mass. Inst. of Technol., Cambridge, March 1974.

Vicens, G. J., I. Rodríguez-Iturbe, and J. C. Schaake, Jr., A Bayesian framework for the use of regional information in hydrology, *Water Resour. Res., 11*(3), 405–414, 1975.

Wood, E. F., I. Rodríguez-Iturbe, and J. C. Schaake, Jr., The methodology of Bayesian inference and decision making applied to extreme hydrologic events, *Rep. 178,* Ralph M. Parsons Lab. for Water Resour. and Hydrodyn., Mass. Inst. of Technol., Cambridge, Jan. 1974.

Zellner, A., *An Introduction to Bayesian Inference in Econometrics,* John Wiley, New York, 1971.

(Received October 21, 1974;
accepted March 21, 1975.)